Methodical settings in analyses of the income distribution: some simple mathematical comments

Faik, Jürgen

Jürgen Faik

Methodical Settings in Analyses of the Income Distribution – Some Simple Mathematical Comments

FaMa-Diskussionspapier 1/2010

FaMa

Neue Frankfurter Sozialforschung

Nikolausstraße 10

D-65936 Frankfurt/Main

Februar 2010

Herausgeber und Verlag:

FaMa

Neue Frankfurter Sozialforschung

Nikolausstraße 10

65936 Frankfurt am Main

Tel.: 069/34409710

Fax: 069/34409714

E-Mail: info@fama-nfs.de

http://www.fama-nfs.de

ISSN 1869-1935

Zusammenfassung*

Das Diskussionspapier befasst sich auf einer teils abstrakten, teils exemplarischen Ebene mit methodischen Fragestellungen im Zusammenhang mit der personellen Einkommensverteilung, um diesbezügliche Grundzusammenhänge offenzulegen, welche dann in einer späteren Arbeit empirisch näher untersucht werden könnten. Es wird ein dekomponierbarer Ungleichheitsindikator aus der Klasse der Generalisierten-Entropie-Indikatoren – der normierte Variationskoeffizient – zugrunde gelegt. Auf dieser Basis werden vor allem die Ungleichheitseinflüsse von Äquivalenzrelationen diskutiert. Hierbei wird in einkommensunabhängige und einkommensabhängige Äquivalenzrelationen unterschieden.

Summary*

The discussion paper deals on a partly abstract, partly exemplary level with methodical issues in the context of the personal income distribution in order to reveal fundamental connections which might be analyzed empirically in a later paper. We will use a decomposable inequality indicator out of the class of Generalized Entropy indicators – the normalized coefficient of variation. On this basis the impacts on inequality, which primarily equivalence relations have, will be discussed. At this, it will be distinguished between income-independent and income-dependent equivalence relations.

* Dr. Jürgen Faik ist Geschäftsführer von FaMa – Neue Frankfurter Sozialforschung. Autoren-Kontakt: faik@fama-nfs.de

Content

1. Introduction

The aim of the paper is to consider some methodical settings in analyses of the income distribution on a rather simple mathematical basis. This shall serve as a mean in order to understand distributional relations better. The content of the paper follows – at least partly – the broad body of literature concerning distributional aspects, which was generated especially in the 1980s (and the early 1990s).[1] Despite this time lag the corresponding questions are yet very important nowadays.

A very crucial setting in distributional studies is the selection of an equivalence scale. It is – to a relatively high degree – a normative decision. So the consequences, following from such a decision, matter strongly in distributional studies. It is the main aim of the paper to show how distributional results may vary with the possibilities for such a selection. For that purpose additively decomposable inequality indicators – like the normalized coefficient of variation – are helpful, particularly in order to distinguish within- from between-subgroup effects which are caused by the influences of different equivalence scales.

In order to reach its sketched main aim the paper is structured as follows: In chapter 2 parameters for distributional sensitivity analyses are considered. After this, in chapter 3 decomposable inequality indicators are presented – especially such out of the class of Generalized Entropy indicators. Chapter 4 deals with such an indicator, the normalized coefficient of variation, and its connection to equivalence relations. We will differentiate between equivalence relations which hold for the whole income range (income-independent equivalence relations), and equivalence relations which distinguish between two (or more) sections of the whole income range (income-dependent equivalence relations). Some concluding remarks in chapter 5 will finish the paper.

2. Parameters for distributional sensitivity considerations

In general, there are a lot of possibilities for dispersing results concerning the personal income distribution due to methodical settings. This includes – in a technical sense – the choice of the inequality indicator, the income definition (or more general: the definition of the used well-being indicator[2]), the selection of the unit of analysis, the length of accounting periods, and the standardizations in consequence of different household sizes and structures.[3]

The characteristics of these variables depend, to some extent, on the concrete issue of research because there are a lot of thinkable specifications of the term “income distribution”. In this context Faik has distinguished the following (general) distributions in the field of personal income distribution:[4]

- Primal versus secondary distribution,
- between- versus within-group distribution,
- between- versus within-generations distribution,
- cross-sectional versus longitudinal distribution, and
- poverty versus richness as margins of the personal distribution.

---

[1] See e. g. the below cited sources Cowell 1980, Shorrocks 1980, Mookherjee/Shorrocks 1982 or Jenkins 1991.
[2] Alternatively, wealth and private consumption could be used as alternative well-being indicators (see e. g. Faik 1995, pp. 36-39).
[3] See e. g. Hussain 2009.
[4] See Faik 2008, pp. 23-24.

So, e. g., in a pre-tax distributional analysis (i. e., in a primal distributional perspective) the gross incomes are relevant, whereas in the context of post-tax distributions (i. e., in a secondary distributional perspective) net income is the central (income) variable. Furthermore – to give another example –, the choice of the temporal horizon is important for the dimension of poverty: Because of possible fluctuations of income during a year a monthly temporal horizon tends to overestimate the degree of poverty compared with a yearly perspective in which such ups and downs would be smoothed.

A more or less general parameter for all of the above mentioned distributional specifications is an equivalence relation. Such relations are necessary in distributional analyses because well-being comparisons between households with different size or (age) structure require a standardization of the used well-being indicator. If household income is taken as such an indicator, the household incomes have to be transformed to household equivalence incomes by dividing the household incomes through so-called equivalence relations. An equivalence relation reflects the economies of scales which arise in bigger households, e. g. because of price or cost advantages compared with smaller households. Additionally, an equivalence relation covers different needs between the household members.[5]

In this context demographic changes can influence income inequality. Empirically, the mean values for the household size decrease in some developed countries like Germany, and there typically exists a positive correlation between household size and household income. The development towards more single-person households tends to reduce the mean of equivalent incomes because a former common household income of a household with two or more persons is now disaggregated into two or more incomes for single-person households without the existence of economies of scale (because in single-person households naturally there are no economies of scale). That means that there seems to be a tendency for reducing the measured income inequality ceteris paribus. But in view of the decreasing mean household sizes in Germany there could be an oppositional effect concerning the inequality of equivalent incomes: The relatively low German fertility rates lead to higher equivalent incomes for couples who decide to have no children (so-called “double income no kids”).[6]

3. Decomposable inequality indicators

3.1 Conceptual framework

For sensitivity analyses the usage of a decomposable inequality indicator is convenient because hereby it is possible to investigate within-group and between-group influences of inequality. The assumed groups must be disjoint to each other.

The within-group component measures the weighted sum of the analyzed indicator for the different groups. Concerning the between-group component each member of a group is given the average income of its group.[7]

[5] Basically see Faik 1995, in the context above especially pp. 39-45.
[6] See Peichl/Pestel/Schneider 2009, p. 2. Furthermore, in a more complex view economically based, indirectly working effects must be considered, e. g. because of ageing processes. Such an incidence analysis requires a more or less detailed economic model (see von Weizsäcker 1988, p. 2).
[7] See e. g. Rodrigues 1993, p. 6.

A very popular class of such indicators is the family of Generalized Entropy (GE) measures (in which the group's population shares serve as weighting factors as well as the income shares of the groups – as we will see later):

(1) $$GE = \frac{1}{(\lambda^2 - \lambda) \cdot n} \cdot \sum_{i=1}^{n} \left[\left(\frac{Y_i}{\mu} \right)^{\lambda} - 1 \right] \qquad \textit{for } \lambda \neq 0 \wedge \lambda \neq 1;$$

$$GE = \frac{1}{n} \cdot \sum_{i=1}^{n} \ln\left(\frac{\mu}{Y_i} \right) \qquad \textit{for } \lambda = 0;$$

$$GE = \frac{1}{n} \cdot \sum_{i=1}^{n} \left[\frac{Y_i}{\mu} \cdot \ln\left(\frac{Y_i}{\mu} \right) \right] \qquad \textit{for } \lambda = 1$$

[GE = General Entropy index, λ = parameter with respect to inequality preferences, n = population size, Y_i = income of person i, μ = mean income].

The parameter λ reflects the social perceptions of inequality. If λ is greater than 0, the upper income area receives a relatively high weight with respect to inequality; the opposite would be the case if λ would be less 0. For $\lambda = 0$ the GE measure represents the mean logarithmic deviation, for $\lambda = 1$ Theil's measure is the result, and for $\lambda = 2$ the GE measure corresponds to the normalized coefficient of variation.

GE can be additively decomposed in a within-group and a between-group component of inequality, as mentioned above:

(2) $$GE = \underbrace{\sum_{g=1}^{G} v_g^{\lambda} \cdot w_g^{1-\lambda} \cdot GE_g}_{\text{within-group inequality}} + \underbrace{GE_B}_{\text{between-group inequality}}$$

The weighting factors w_g (= n_g/n) represent the population shares of the several groups of persons g (g = 1, 2, …, G), μ_g is the mean of incomes within group g, v_g (= $w_g\ \mu_g/\mu$) denotes the group-specific share of the aggregate income, and GE_g symbolizes the within-group GE inequality measure and GE_B the between-group GE inequality indicator.

At this, GE_B is defined in the following way:

(3) $$GE_B = \frac{1}{(\lambda^2 - \lambda)} \cdot \left\{ \left[\sum_{g=1}^{G} w_g \cdot \left(\frac{\mu_g}{\mu} \right)^{\lambda} \right] - 1 \right\} \qquad \textit{for } \lambda \neq 0 \wedge \lambda \neq 1;$$

$$GE_B = \sum_{g=1}^{G} w_g \cdot \ln\left(\frac{\mu}{\mu_g} \right) \qquad \textit{for } \lambda = 0;$$

$$GE_{\mathrm{B}} = \sum_{g=1}^{G} v_g \cdot \ln\left(\frac{\mu_g}{\mu}\right) \quad \textit{for } \lambda = 1.$$ [8]

In the following the three above mentioned GE indicators – the mean logarithmic deviation, Theil's indicator of entropy, and the normalized coefficient of variation – are discussed. This discussion includes a detailed definition of each of these indicators as well as a consideration concerning the decomposition for each of the three measures.

3.2 Mean logarithmic deviation

3.2.1 Definition

A good approximation for empirical income distributions, which are typically right-skewed, is the log-normal distribution. One of its parameters is the standard deviation of the logarithms of income (L_1). Because of this, it is not far to seek that this distributional parameter is an intuitively plausible indicator for the measurement of income inequality. Its definition is as follows:

$$(4) \quad L_1 = \sqrt[2]{\frac{1}{n} \cdot \sum_{i=1}^{n} (\ln Y_i - \ln \phi)^2}$$

$$= \sqrt[2]{\frac{1}{n} \cdot \sum_{i=1}^{n} \ln\left(\frac{Y_i}{\phi}\right)^2}$$

[with: L_1 = standard deviation of the logarithms of income; ϕ = geometric mean of incomes: $\ln\phi = \frac{1}{n} \cdot \sum_{i=1}^{n} \ln Y_i$].

Sometimes L_1 is approximated by the logarithmic standard deviation (L_2):

$$(5) \quad L_2 = \sqrt[2]{\frac{1}{n} \cdot \sum_{i=1}^{n} (\ln Y_i - \ln \mu)^2}$$

$$= \sqrt[2]{\frac{1}{n} \cdot \sum_{i=1}^{n} \ln\left(\frac{Y_i}{\mu}\right)^2}$$

[with μ as arithmetic mean of incomes].[9]

A similar measure – compared with L_2 – is the mean logarithmic deviation which results in the GE family at $\lambda = 0$ (see formula (1)). In contrast to L_2, the latter mentioned measure is not based on squared but only on simple deviations of the Y_i values from μ.

---

[8] A more comprehensive consideration of the class of GE measures can be found in Faik 1995, pp. 326-330, which is primarily based on Cowell 1980, Shorrocks 1980, Mookherjee/Shorrocks 1982 and Jenkins 1991.

[9] See Faik 1995, pp. 301-302.

3.2.2 Decomposition

That the mean logarithmic deviation (MLD) can be decomposed was demonstrated by Mookherjee/Shorrocks.[10] Under the prerequisite that the whole population can be differentiated into G disjoint subgroups we obtain:

(6) $$MLD = \sum_{g=1}^{G} \frac{n_g}{n} \cdot MLD_g + \sum_{g=1}^{G} \frac{n_g}{n} \cdot ln\left(\frac{\mu}{\mu_g}\right)$$

$$\Leftrightarrow MLD = \underbrace{\sum_{g=1}^{G} w_g \cdot MLD_g}_{\text{within-group inequality}} + \underbrace{\sum_{g=1}^{G} w_g \cdot ln\left(\frac{\mu}{\mu_g}\right)}_{\text{between-group inequality}}.$$

Furthermore, the variation of MLD over time in the sense of the difference between MLD at the time periods t+1 and t can be decomposed into four terms (a bar over variables denotes the average value of t and t+1 values):

(7) $$\Delta MLD = MLD_{t+1} - MLD_t$$

$$\Leftrightarrow \Delta MLD = \sum_{g=1}^{G} \overline{w_g} \cdot \Delta MLD_g + \sum_{g=1}^{G} \overline{MLD} \cdot \Delta w_g - \sum_{g=1}^{G} \overline{ln\left(\frac{\mu_g}{\mu}\right)} \cdot \Delta w_g - \sum_{g=1}^{G} \overline{w_g} \cdot \Delta\, ln\left(\frac{\mu_g}{\mu}\right)$$

$$\Leftrightarrow \Delta MLD = \sum_{g=1}^{G} \overline{w_g} \cdot \Delta MLD_g \quad (term\ \ A) + \sum_{g=1}^{G} \overline{MLD_g} \cdot \Delta w_g \quad (term\ \ B)$$

$$- \sum_{g=1}^{G} \left[\overline{\left(\frac{\mu_g}{\mu}\right)} \cdot \overline{ln\left(\frac{\mu_g}{\mu}\right)}\right] \cdot \Delta w_g \quad (term\ \ C) - \sum_{g=1}^{G} \left[\overline{\left(w_g \cdot \frac{\mu_g}{\mu}\right)} - \overline{w_g}\right] \cdot \Delta\, ln\left(\mu_g\right) \ (term\ \ D).$$

Term A reflects the impact of intertemporal changes concerning the within-group inequality, term B represents changes in the within-group inequality due to changes in the population shares, term C indicates the impact of such changes in the population shares upon the between-group inequality, and term D characterizes relative changes in the group's mean incomes and their influence upon the overall inequality changes.[11]

---

[10] See Mookherjee/Shorrocks 1982, pp. 896-897.

[11] See originally Mookherjee/Shorrocks 1982; see Rodrigues 1993, p. 9, or Peichl/Pestel/Schneider 2009, pp. 7-9, too. In the latter paper an empirical application of the above formula for Germany (on the basis of the Socioeconomic Panel) can be found.

Summarizing the afore-mentioned characterizations,

- Term A is an expression for the within-group inequality (generated by differing characteristics within the groups which are others than the group-constituting characteristics),
- the terms B and C represent the demographical component of the inequality change, and
- term D reflects the impact of changes in the distribution of average incomes upon the several groups.[12]

3.3 Theil's indicator of entropy

3.3.1 Definition

In the context of Theil's indicator of entropy there exist analogies to the information theory. Starting point is the consideration that an event is as much more interestingly, when its probability of occurrence ψ is low, e. g. depicted in the following way:

(8) $$h(\psi) = \ln \frac{1}{\psi}.$$

Summing the n single events to an overall event leads to the expected value:

(9) $$H(\psi) = \sum_{i=1}^{n} \psi_i \cdot h(\psi_i)$$

$$= \sum_{i=1}^{n} \psi_i \cdot \ln \frac{1}{\psi_i}; \quad \text{with:} \quad \sum_{i=1}^{n} \psi_i = 1.$$

Equation (9) has its maximum at $\psi_i = 1/n$ (for i = 1, 2, …, n); this maximum H_{max} is called the maximal entropy:

(10) $$H_{max}(\psi) = \sum_{i=1}^{n} \psi_i \cdot \ln \left[\frac{1}{\frac{1}{n}} \right]$$

$$= \ln n \cdot \sum_{i=1}^{n} \psi_i$$

$$= \ln n.$$

In considerations of income inequality the ψ_i reflect the income shares $\frac{Y_i}{\sum_{i=1}^{n} Y_i}$.

[12] See Peichl/Pestel/Schneider 2009, pp. 8-9.

If the income shares are distributed entirely equally, so that the income share of each of the n units of analysis amounts to 1/n, obviously H_{max} is reached. This case should be depicted by a value of 0 for the final entropy measure; in order to realize that, the difference between the maximal and the actual entropy of the distribution must be computed. In that way Theil's indicator of entropy (T) is defined by:

(11) $$T = H_{max}(\Psi) - H(\Psi)$$

$$\Leftrightarrow T = \ln n - \sum_{i=1}^{n} \psi_i \cdot \ln \frac{1}{\psi_i}$$

$$\Leftrightarrow T = \frac{1}{n \cdot \mu} \cdot \left[\sum_{i=1}^{n} (Y_i \cdot \ln Y_i) - \ln \mu \cdot \sum_{i=1}^{n} Y_i \right]$$

$$\Leftrightarrow T = \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{Y_i}{\mu} \cdot \ln \left(\frac{Y_i}{\mu} \right).$$

[with: T = Theil's indicator of entropy; $H_{max}(\psi)$ = maximal entropy; $H(\psi)$ = entropy of the distribution].

As outlined above, T has its minimal value in the amount of zero (when a uniform distribution exists); its maximum value is – in the case of a completely unequal distribution – ln n which can be shown in a consideration of limit values (with H equals ln (1) = 0).

3.3.2 Decomposition

Starting from the final formula of expression (11), i. e. starting from the non-decomposed formula of Theil's measure, we have to distinguish between G subgroups with n_g members respectively. The corresponding income values within each group g (g = 1, 2, ..., G) are named by $Y_{i,g}$, so that – keeping in mind the symmetry of the ranking of incomes for T – expression (11) can be transformed to (the bold symbols represent vectors):

(12) $$T = T(\mathbf{Y}, n)$$

$$= T(\mathbf{Y_1}, \mathbf{Y_2}, ..., \mathbf{Y_G}; n)$$

$$= \frac{1}{n} \cdot \sum_{g=1}^{G} \sum_{i=1}^{n_g} \frac{Y_{i,g}}{\mu} \cdot \ln \left(\frac{Y_{i,g}}{\mu} \right).$$ [13]

[13] See Shorrocks 1980, p. 613.

In the next step the right side of (12) is extended by $[(\mu_g\ n_g)/(\mu_g\ n_g)]$, and the logarithmic expression by (μ_g/μ_g). Subsequently, some terms are substituted by others, and some expressions are multiplied out, so that the following results:

$$(13)\quad T = \sum_{g=1}^{G}\left(\frac{\mu_g \cdot n_g}{\mu \cdot n}\right)\cdot\frac{1}{n_g}\cdot\left\{\sum_{i=1}^{n_g}\left(\frac{Y_{i,g}}{\mu_g}\cdot ln\frac{Y_{i,g}}{\mu_g}\right)+\sum_{i=1}^{n_g}\left(\frac{Y_{i,g}}{\mu_g}\cdot ln\frac{\mu_g}{\mu}\right)\right\}$$

$$= \sum_{g=1}^{G}\left(\frac{\mu_g \cdot n_g}{\mu \cdot n}\right)\cdot\frac{1}{n_g}\cdot\sum_{i=1}^{n_g}\left(\frac{Y_{i,g}}{\mu_g}\cdot ln\frac{Y_{i,g}}{\mu_g}\right)$$

$$+ \sum_{g=1}^{G}\left(\frac{\mu_g \cdot n_g}{\mu \cdot n}\right)\cdot\frac{1}{n_g}\cdot\sum_{i=1}^{n_g}\left(\frac{Y_{i,g}}{\mu_g}\cdot ln\frac{\mu_g}{\mu}\right).$$

Considering the group-specific measure of T

$$(14)\quad T_g = \frac{1}{n_g}\cdot\sum_{i=1}^{n_g}\left[\frac{Y_{i,g}}{\mu_g}\cdot ln\left(\frac{Y_{i,g}}{\mu_g}\right)\right],$$

it results:

$$(15)\quad T = \sum_{g=1}^{G}\left(\frac{\mu_g \cdot n_g}{\mu \cdot n}\right)\cdot T_g + \sum_{g=1}^{G}\left(\frac{\mu_g \cdot n_g}{\mu \cdot n}\right)\cdot\frac{1}{n_g}\cdot\sum_{i=1}^{n_g}\left[\frac{Y_{i,g}}{\mu_g}\cdot ln\left(\frac{\mu_g}{\mu}\right)\right],$$

and thus:

$$(16)\quad T = \sum_{g=1}^{G}\left(\frac{\mu_g \cdot n_g}{\mu \cdot n}\right)\cdot T_g + \sum_{g=1}^{G}\left(\frac{\mu_g \cdot n_g}{\mu \cdot n}\right)\cdot\frac{1}{\mu_g}\cdot ln\left(\frac{\mu_g}{\mu}\right)\cdot\frac{1}{n_g}\cdot\sum_{i=1}^{n_g}Y_{i,g}\ .$$

With the group-specific arithmetic mean

$$(17)\quad \mu_g = \frac{1}{n_g}\cdot\sum_{i=1}^{n_g}Y_{i,g}\ ,$$

last but not least we obtain – after the mathematical operation of cancelling of some expressions – Theil's additively decomposed indicator:

$$(18)\quad T = \underbrace{\sum_{g=1}^{G}v_g \cdot T_g}_{\text{within-group inequality}} + \underbrace{\sum_{g=1}^{G}v_g \cdot ln\left(\frac{\mu_g}{\mu}\right)}_{\text{between-group inequality}}$$

[with: $v_g = (n_g\ \mu_g)/(n\ \mu)$].[14]

[14] See Shorrocks 1980, p. 613, too.

3.4 Normalized coefficient of variation

3.4.1 Definition

An inequality indicator which refers to the first two moments of a distribution, is the coefficient of variation (V):[15]

(19) $V = \frac{S}{\mu}$

[with: V = coefficient of variation, S = standard deviation of incomes, μ = arithmetic mean of incomes].

The so-called normalized coefficient of variation (CV), which – as mentioned above – belongs to the class of GE measures (at λ = 2), is a simple transformation of (19):

(20) $CV = \frac{1}{2} \cdot \frac{S^2}{\mu^2}$.

In the following we will show that CV can be interpreted as a GE measure. In order to reach this aim, we specify the above definition of a GE measure (see equation (1)) by fixing λ at the value of 2:

(21) $GE^{(2)} = \frac{1}{(2^2 - 2) \cdot n} \cdot \sum_{i=1}^{n} \left[\left(\frac{Y_i}{\mu} \right)^2 - 1 \right]$

$$\Leftrightarrow GE^{(2)} = \frac{1}{2 \cdot n} \cdot \sum_{i=1}^{n} \left[\left(\frac{Y_i}{\mu} \right)^2 - 1 \right]$$

$$\Leftrightarrow GE^{(2)} = \frac{1}{2 \cdot n} \cdot \sum_{i=1}^{n} \left[\frac{Y_i^2}{\mu^2} - 1 \right]$$

$$\Leftrightarrow GE^{(2)} = \frac{1}{2 \cdot n} \cdot \sum_{i=1}^{n} \left[\frac{Y_i^2 - \mu^2}{\mu^2} \right]$$

$$\Leftrightarrow GE^{(2)} = \frac{1}{2 \cdot n} \cdot \frac{1}{\mu^2} \cdot \sum_{i=1}^{n} \left[Y_i^2 - \mu^2 \right]$$

$$\Leftrightarrow GE^{(2)} = \frac{1}{2} \cdot \frac{1}{\mu^2} \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} \left[Y_i^2 - \mu^2 \right]$$

$$\Leftrightarrow GE^{(2)} = \frac{1}{2} \cdot \frac{1}{\mu^2} \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} Y_i^2 - \frac{1}{2} \cdot \frac{1}{\mu^2} \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} \mu^2$$

$$\Leftrightarrow GE^{(2)} = \frac{1}{2} \cdot \frac{1}{\mu^2} \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} Y_i^2 - \frac{1}{2} \cdot \frac{1}{\mu^2} \cdot \frac{1}{n} \cdot n \cdot \mu^2$$

$$\Leftrightarrow GE^{(2)} = \frac{1}{2} \cdot \frac{1}{\mu^2} \cdot \left[\left(\frac{1}{n} \cdot \sum_{i=1}^{n} Y_i^2 \right) - \mu^2 \right]$$

---

[15] See e. g. Faik 2007, pp. 95-96.

$$\Leftrightarrow GE^{(2)} = \frac{1}{2}\cdot\frac{1}{\mu^2}\cdot S^2 = \frac{1}{2}\cdot\frac{S^2}{\mu^2} \equiv CV \quad \text{(q. e. d.)}$$

3.4.2 Decomposition

The decomposition of CV in the context of a GE measure can be shown as follows (in a backward procedure of computation):

(22)
$$GE^{(2)} = \sum_{g=1}^{G} v_g^2 \cdot w_g^{-1} \cdot GE_g^{(2)} + \frac{1}{2}\cdot\left\{\left[\sum_{g=1}^{G} w_g \cdot \left(\frac{\mu_g}{\mu}\right)^2\right] - 1\right\}$$

$$with: \quad w_g = \frac{n_g}{n},\; v_g = w_g \cdot \frac{\mu_g}{\mu}$$

$$\Leftrightarrow GE^{(2)} = \sum_{g=1}^{G} \frac{n_g^2}{n^2}\cdot\frac{\mu_g^2}{\mu^2}\cdot\frac{n}{n_g}\cdot\frac{1}{2}\cdot\frac{S_g^2}{\mu_g^2} + \frac{1}{2}\cdot\left\{\left[\sum_{g=1}^{G}\frac{n_g}{n}\cdot\frac{\mu_g^2}{\mu^2}\right] - 1\right\}$$

$$\Leftrightarrow GE^{(2)} = \sum_{g=1}^{G} \frac{n_g^2}{n^2}\cdot\frac{1}{\mu^2}\cdot\frac{1}{2}\cdot S_g^2 + \frac{1}{2}\cdot\sum_{g=1}^{G}\frac{n_g}{n}\cdot\frac{\mu_g^2}{\mu^2} - \frac{1}{2}$$

$$\Leftrightarrow GE^{(2)} = \frac{1}{2}\cdot\frac{\sum_{g=1}^{G}\frac{n_g}{n}\cdot S_g^2 + \sum_{g=1}^{G}\frac{n_g}{n}\cdot\mu_g^2 - \mu^2}{\mu^2}$$

$$\Leftrightarrow GE^{(2)} = \frac{1}{2}\cdot\frac{\sum_{g=1}^{G}\frac{n_g}{n}\cdot\left[\left(\frac{1}{n_g}\cdot\sum_{i=1}^{n_g}Y_{i,g}^2\right) - \mu_g^2\right] + \sum_{g=1}^{G}\frac{n_g}{n}\cdot\mu_g^2 - \mu^2}{\mu^2}$$

$$\Leftrightarrow GE^{(2)} = \frac{1}{2}\cdot\frac{\sum_{g=1}^{G}\frac{n_g}{n}\cdot\frac{1}{n_g}\cdot\sum_{i=1}^{n_g}Y_{i,g}^2 - \sum_{g=1}^{G}\frac{n_g}{n}\cdot\mu_g^2 + \sum_{g=1}^{G}\frac{n_g}{n}\cdot\mu_g^2 - \mu^2}{\mu^2}$$

$$\Leftrightarrow GE^{(2)} = \frac{1}{2}\cdot\frac{\frac{1}{n}\sum_{g=1}^{G}\sum_{i=1}^{n_g}Y_{i,g}^2 - \mu^2}{\mu^2}$$

$$\Leftrightarrow GE^{(2)} = \frac{1}{2}\cdot\frac{S^2}{\mu^2} \equiv CV \quad \text{(q. e. d.)}$$

4. The coefficient of variation and equivalence relations

4.1 Equivalence relations across the whole income range

4.1.1 Fundamental connections

In the context of inequality of equivalent incomes the correlation between household size and household income matters. Typically, there is a positive correlation between these two variables. Starting with the assumption of highest economies of scales and, thus, equivalence relations in the amount of 1.0 for all household types, subsequently the degree of economies of scales is reduced stepwise corresponding to higher equivalence relations which means a levelling concerning the equivalent household incomes. Shortly spoken: The measured inequality decreases. But the further dropping of the bigger household's equivalent incomes will lead to an increase in the measured inequality at a certain point. So a u-shaped curve for the inequality levels depending on the range of economies of scales is realistic.

If a negative correlation between household size and household income occurs, it is probable that the inequality curve has a positive slope across the whole area or most of the area of scale values which begins with the "per household situation" and ends with the "per capita situation". In this case and in an ideal-typical perspective, the relatively low incomes of the bigger household sizes – compared with the smaller household sizes – would be reduced continuously, and so the inequality between the different household sizes would arise.

Formally, for CV the influence of equivalence relations on the measured inequality can be shown by substituting the household incomes within the g groups (g = 1, 2, ..., G) $Y_{i,g}$ through the equivalent incomes within the several groups $Y_{i,g}/m_g$. In order to detect the influence of the equivalence relations m_g on the measured inequality (by CV), we can decompose the two main elements of CV, namely μ and S^2 (the asterisk denotes equivalent income):

(23) $$\mu_* = \sum_{g=1}^{G} \frac{\mu_g}{m_g} \cdot \frac{n_g}{n},$$

$$S^2_{*,within} = \sum_{g=1}^{G} \frac{S_g^2}{m_g^2} \cdot \frac{n_g}{n},$$

$$S^2_{*,between} = \sum_{g=1}^{G} \left(\frac{\mu_g}{m_g} - \mu_* \right)^2 \cdot \frac{n_g}{n}$$

$$\Rightarrow CV = \frac{1}{2} \cdot \frac{\sum_{g=1}^{G} \frac{S_g^2}{m_g^2} \cdot \frac{n_g}{n} + \sum_{g=1}^{G} \left(\frac{\mu_g}{m_g} - \mu_* \right)^2 \cdot \frac{n_g}{n}}{\left(\sum_{g=1}^{G} \frac{\mu_g}{m_g} \cdot \frac{n_g}{n} \right)^2}.$$

Whereas the overall values of CV change with variations of the equivalence relations, the within-group CVs do not change with diminishing economies of scale. The denominator of CV_g (the squared arithmetic mean of equivalent incomes within group g) is multiplied by the same factor – the inverse of the squared equivalence relation which holds for group g – as the nominator of CV_g (the variance of equivalent incomes within group g). So, these factors will be cancelled by computing CV_g, and the within-group CVs remain constantly across the range of economies of scale (from "complete economies of scale" up to "no economies of scale at all"). This is intuitively plausible because for the within-group distribution of incomes equivalence relations do not matter. This is because the underlying units of analysis are homogenous within the groups (at least concerning the group-constituting characteristics).

This can be simply shown in the following way, at first for the denominator of CV_g (the squared arithmetic mean; the asterisk denotes the fact that the incomes are deflated by equivalence relations):

$$(24)\quad \left(\mu_g^2\right)_* = \left(\frac{1}{n}\sum_{i=1}^{n_g}\frac{Y_{i,g}}{m_g}\right)^2$$

$$\Leftrightarrow \left(\mu_g^2\right)_* = \frac{1}{m_g^2}\left(\frac{1}{n}\sum_{i=1}^{n_g}Y_{i,g}\right)^2$$

$$\Leftrightarrow \left(\mu_g^2\right)_* = \frac{1}{m_g^2}\cdot\mu^2 .$$

For the numerator of CV_g (the variance of equivalent incomes) the mathematical proof is as follows:

$$(25)\quad \left(S_g^2\right)_* = \frac{1}{n}\sum_{i=1}^{n_g}\frac{Y_{i,g}^2}{m_g^2} - \frac{1}{m_g^2}\cdot\mu^2$$

$$\Leftrightarrow \left(S_g^2\right)_* = \frac{1}{m_g^2}\cdot\left[\frac{1}{n}\cdot\sum_{i=1}^{n_g}Y_{i,g}^2 - \mu^2\right]$$

$$\Leftrightarrow \left(S_g^2\right)_* = \frac{1}{m_g^2}\cdot S^2 .$$

4.1.2 Examples

The foregoing statements will be illustrated by some examples. In the first example we differentiate between a positive correlation of household size versus (non-adjusted) household income on the one hand and a negative correlation of the two stated variables on the other hand. The equivalence relations are varied by the formula $m_g = g^\theta$ in which g marks the household size and θ denotes in a well-known way the economies of scale.[16]

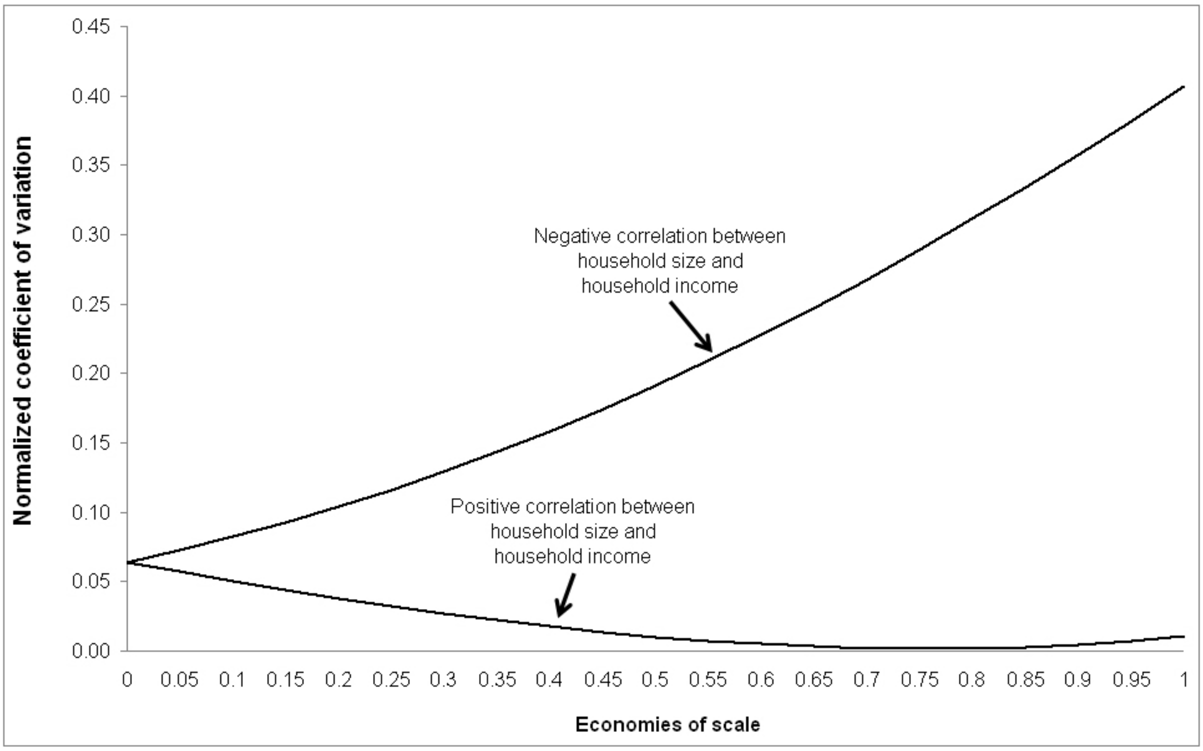
Example 1:

Household size	Household income (case A: positive correlation)	Household income (case B: negative correlation)
1 person	1,000	3,500
2 persons	2,000	3,000
3 persons	2,500	2,500
4 persons	3,000	2,000
5 persons	3,500	1,000

It becomes evident, as can be seen in figure 1, that in the case of a positive correlation a u-shaped function of CV results with a minimum at $\theta = 0.75$, whereas in the case of a negative correlation a continuously positive slope of the inequality function was generated. These results exemplify the above statements.

[16] The underlying equivalence scale is the so-called Buhmann et al. scale (see Buhmann et al. 1988, p. 119). For a comprehensive discussion of this scale see Faik 2009, p. 6.

Figure 1: Functional form of the normalized coefficient of variation
in dependence of decreasing economies of scale
(Buhmann et al. approach; example 1)



Source: Own illustration

In the following we will give some further examples in order to check the sensitivity of CV. Concretely, for two groups we will vary the following parameters:

- The relation of the within-group CVs: Equality versus inequality,
- the population shares,
- once more the type of the correlation between household size and household income: Positive versus negative correlation, and
- the kind of equivalence relations: Buhmann et al. versus Citro/Michael approach.

All these simulations are compiled in the following table.

Example 2:

Household type	Equiva-lence relation	Mean income (in a mone-tary unit)	Standard deviation of incomes (in a mone-tary unit)	Population share	Normalized coefficient of variation within the group
Example 2A:					
1 person	1.0^{θ}	1,000	500	0.40	0.125
2 persons	2.0^{θ}	1,500	750	0.60	0.125
Example 2B:					
1 person	1.0^{θ}	1,000	800	0.40	0.320
2 persons	2.0^{θ}	1,500	200	0.60	0.009
Example 2C:					
1 person	1.0^{θ}	1,000	200	0.40	0.020
2 persons	2.0^{θ}	1,500	1,000	0.60	0.222
Example 2D:					
1 person	1.0^{θ}	1,000	500	0.10/.../0.90	0.125
2 persons	2.0^{θ}	1,500	750	0.90/.../0.10	0.125
Example 2E:					
1 person	1.0^{θ}	1,500	750	0.40	0.125
2 persons	2.0^{θ}	1,000	500	0.60	0.125
Example 2F:					
1 person	1.0^{θ}	1,000	500	0.40	0.125
2 persons	1.7^{θ}	1,500	750	0.60	0.125

The general formula for computing CV in all these simulations is (based on formula (22)):

(26) $CV = CV_I + CV_{II}$

with: $$CV_I = \frac{n_1}{n} \cdot \left(\frac{\mu_1}{\mu_*}\right)^2 \cdot CV_1 + \frac{n_2}{n} \cdot \left(\frac{\frac{\mu_2}{m_2}}{\mu_*}\right)^2 \cdot CV_2 ,$$

$$CV_{II} = \left[\left\{\frac{n_1}{n} \cdot \left(\frac{\mu_1}{\mu_*}\right)^2 + \frac{n_2}{n} \cdot \left(\frac{\frac{\mu_2}{m_2}}{\mu_*}\right)^2\right\} - 1\right] \cdot 0.5 \text{, and}$$

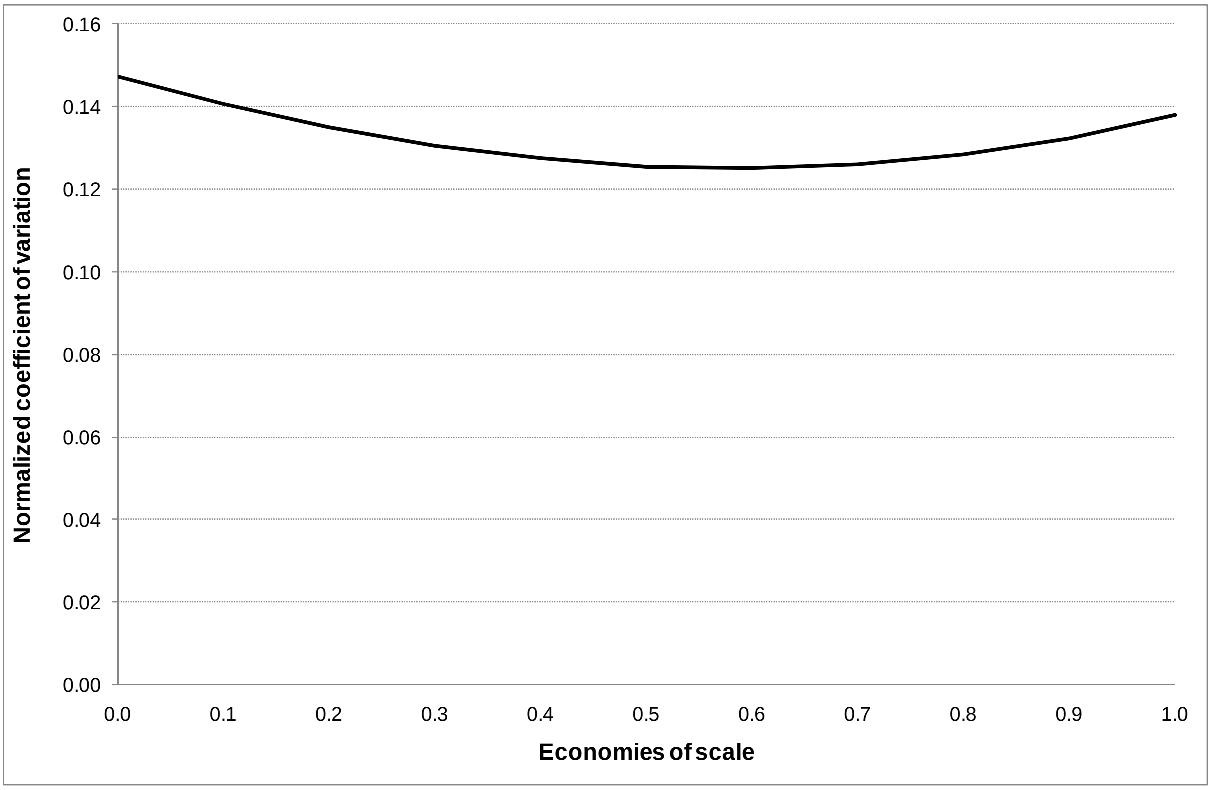
$$\mu_* = \frac{n_1}{n} \cdot \mu_1 + \frac{n_2}{n} \cdot \frac{\mu_2}{m_2}$$

[with: CV_I = within-group element of CV; CV_{II} = between-group element of CV].

Figure 2 illustrates as a starting point for the further sensitivity considerations example 2A in which the Buhmann et al. approach for creating equivalence relations is assumed as well as a positive correlation between household size and household income. Furthermore, equal normalized coefficients of variation within the two groups are used. The assumed relation of the population shares between single-person and two-person households is 40 per cent to

60 per cent. As a result we obtain the u-shaped inequality curve for the overall coefficient of variation.

Figure 2: Functional form of the normalized coefficient of variation (CV)
in dependence of decreasing economies of scale for two groups of persons:
Positive correlation between household size and household income,
equal within-group CVs (Buhmann et al. approach; example 2A)



Source: Own illustration

Compared with the reference curve presented in figure 2, the variation of the within-group CVs in example 2B causes for a higher value of CV within the single-person households than within the two-person households a positive slope of the inequality curve (see figure 3); at θ = 0.50 there is a point of intersection with the reference curve in which the same CV values between single-person and two-person households were assumed. In contrast, the inequality curve caused by example 2C is negatively sloped with a point of intersection with the reference curve at θ = 0.80.[17]

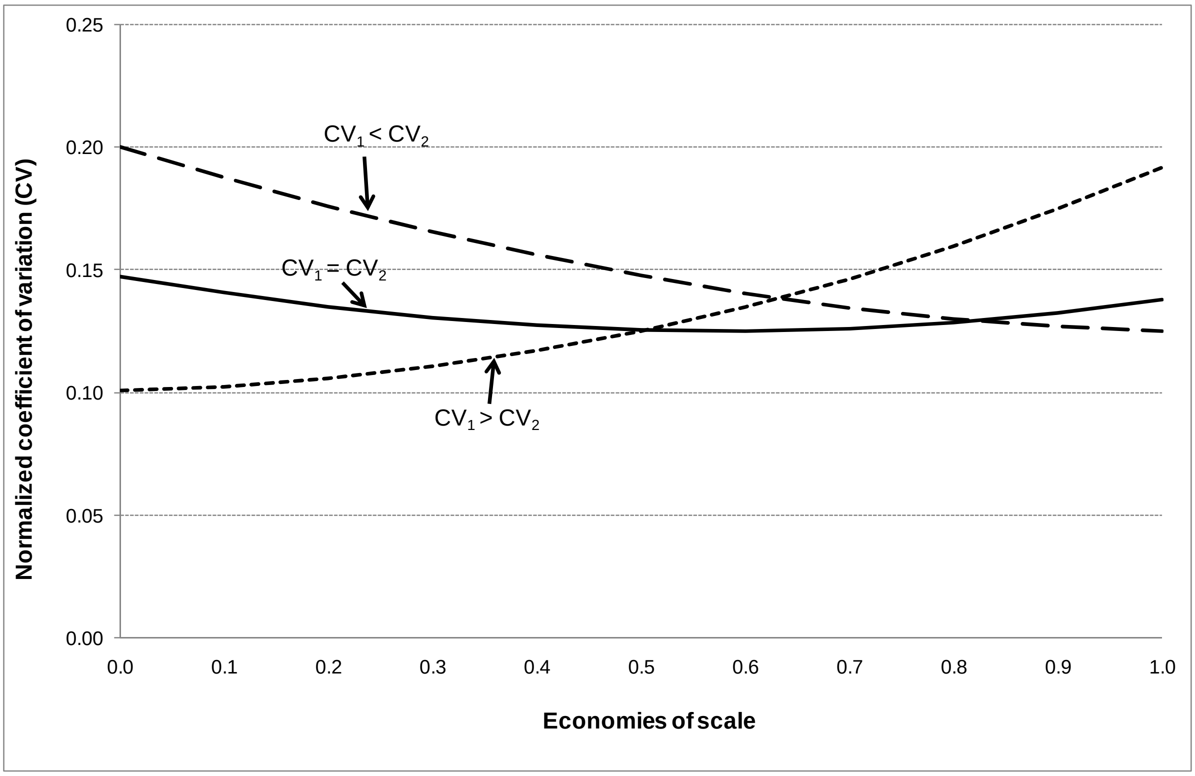
In example 2B, when CV_1 (i. e., the normalized coefficient of variation for the single-person households) is greater than CV_2 (i. e., the normalized coefficient of variation for the two-person households) the first term of CV_I in equation (26) has a higher weight (represented by CV_1) than in example 2A. Furthermore, the overall mean of equivalent income μ_* decreases with higher values of the equivalence relations (which mean: with a lower degree of economies of scale). This has – ceteris paribus – an increasing effect on the overall value of CV through an increase of the first terms of CV_I and CV_{II} whereas the effect on the second terms of CV_I and CV_{II} is an opposite one because the diminishing effect through μ_* is counteracted through rising equivalence relations for the two-person households only in an under propor-

[17] By the way, in the examples 2A-2C a variation of the population shares would only change the locations of the curves but not their forms, as further – here not presented – own simulations have shown.

tional way which causes – ceteris paribus – a diminishing effect on measured inequality. The latter effect is weighted by a relatively low value of CV_2 whereas the first of all stated effects – i. e. the increasing inequality due to the first terms of CV_I and CV_{II} – has a relatively high weight (in the form of CV_1). All in all, the last-mentioned effect dominates the other (i. e., the opposite) effect so that a positive slope of the inequality curve is the consequence, and the overall inequality rises with increasing values of θ.

The opposite is the case in example 2C: Here the relevance of the sketched effect of increasing inequality (for single-person households) is weakened by a relatively low value of the weight CV_1, and the effect of decreasing inequality (for two-person households) is strengthened by a relatively high value of the corresponding weight CV_2 so that as a consequence the overall inequality decreases with higher values of θ.

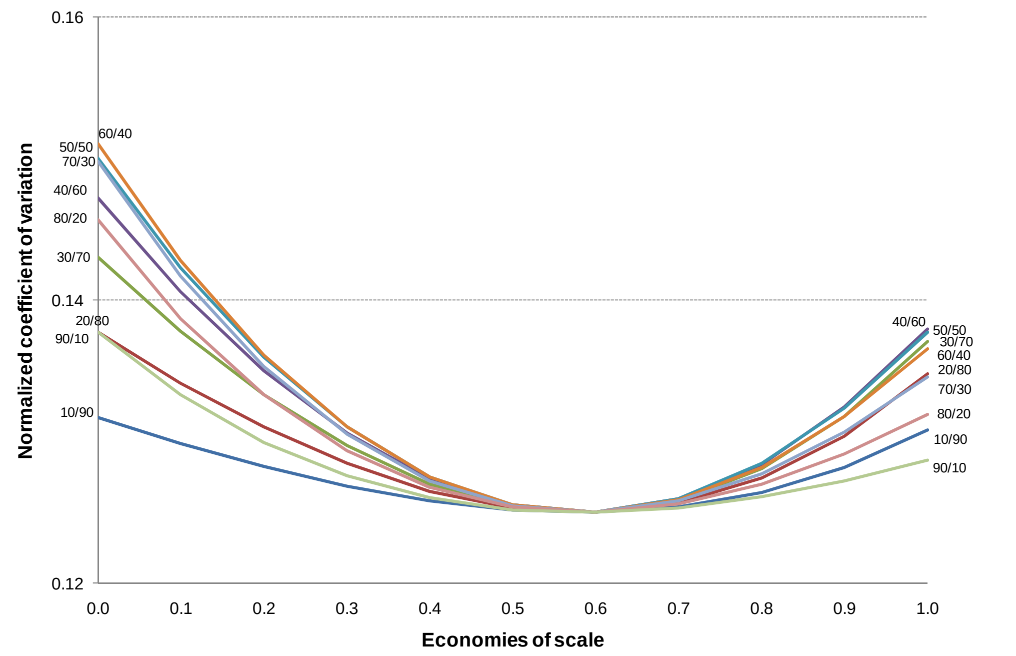
Figure 3: Functional form of the normalized coefficient of variation (CV)
in dependence of decreasing economies of scale for two groups of persons:
Positive correlation between household size and household income,
equal versus unequal within-group CVs
(Buhmann et al. approach; examples 2B and 2C)



Source: Own illustration

Figure 4 on the next page illustrates example 2D in which the population shares of the compared household types (single-person versus two-person households) are varied. In order to show the differences between the different variants distinctly, the ordinate starts with the value 0.12. Otherwise – i. e. with a starting value for CV in the amount of zero – it would be difficult to detect any differences between the several variants with the naked eye. So, firstly, we can register that in our example no big differences between the stated variants exist, and, secondly, there is no clear, no systematic ordering of the different variants across the area of θ.

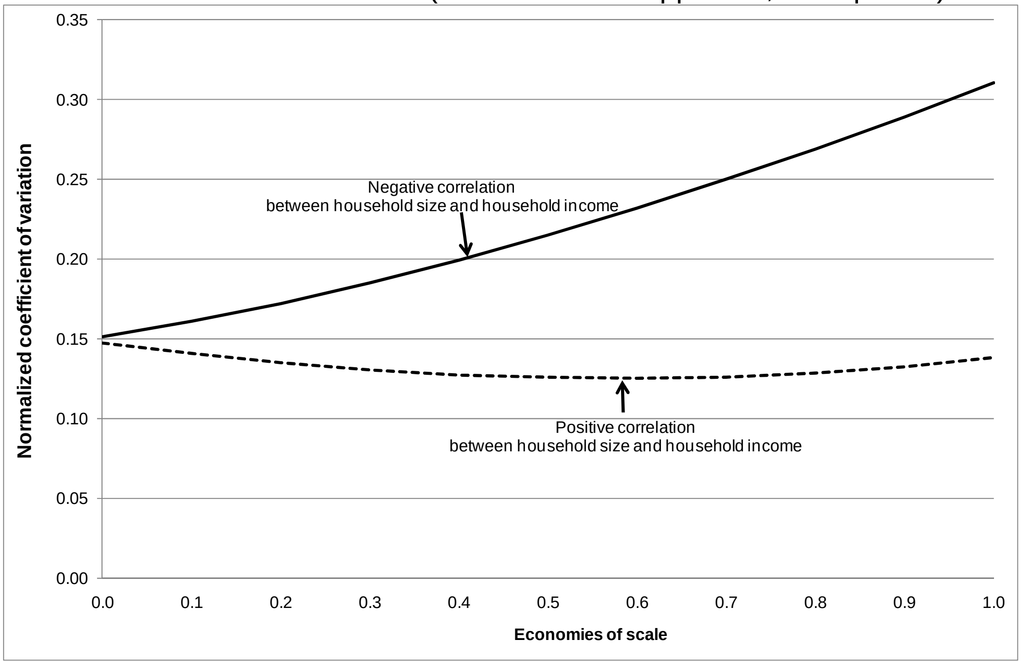
Figure 4: Functional form of the normalized coefficient of variation (CV)
in dependence of decreasing economies of scale for two groups of persons:
Positive correlation between household size and household income,
varying population shares (in per cent; Buhmann et al. approach; example 2D)



Source: Own illustration

Quasi for the sake of completeness, figure 5 shows the differences between example 2 E and example 2A: The results of the above example 1 are qualitatively reflected which is not all too much surprisingly.

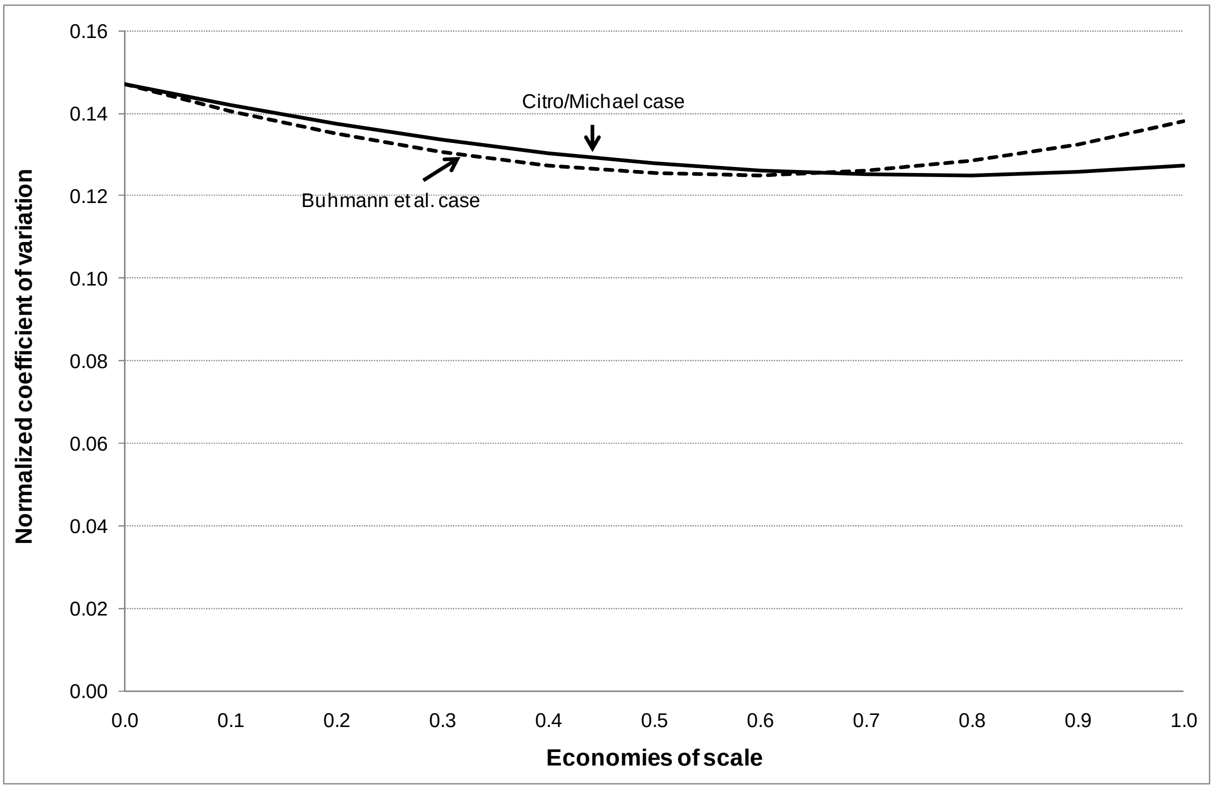
Figure 5: Functional form of the normalized coefficient of variation (CV)
in dependence of decreasing economies of scale for two groups of persons:
Positive versus negative correlation between household size
and household income (Buhmann et al. approach; example 2E)



Source: Own illustration

In addition to the preceding examples, in figure 6 the differences between the Buhmann et al. and the Citro/Michael approach for constituting equivalence relations are presented. Citro/Michael[18] have split the Buhmann et al. approach insofar into two parts as they have differentiated household size into the number of adults and into the number of children. In our example 2F we assume for reasons of simplicity that all two-person households represent households consisting of one adult and one child. In this context a child shall be equivalent to 0.7 adults (i. e. needs for children which are 30 per cent less than that for adults). With this specification we obtain until $\theta \approx 0.65$ slight higher inequality values and from this point lower inequality values in the Citro/Michael approach compared with the Buhmann et al. formula. As can be seen by formula (26), the influence of the two different equivalence scales was partly constituted by the definition of the equivalent income mean μ_* which in our example was increased by $\frac{\left(2^{\theta}-1.7^{\theta}\right)}{2^{\theta}\cdot 1.7^{\theta}}\cdot\frac{n_2}{n}\cdot\mu_2$ in the Citro/Michael case compared with the Buhmann et al. case. This causes a diminishing effect with respect to the contributions of the single-person households in the first terms of CV_I and CV_{II} and so ceteris paribus a reduction of CV. But this effect is counteracted to some degree by the contributions of the two-person households in the second terms of CV_I and CV_{II}. Although the effect of μ_*, which reduces the inequality, exists, too, there is a reverse effect – upon CV through the second terms of CV_I and CV_{II} – generated by the lower equivalence relations in the Citro/Michael approach compared with the Buhmann et al. formula. Obviously, in figure 6 this latter effect dominates the first sketched effects of μ_* up to $\theta \approx 0.65$, and from this point the contrary is the case.

Figure 6: Functional form of the normalized coefficient of variation (CV) in dependence of decreasing economies of scale for two groups of persons: Buhmann et al. versus Citro/Michael approach (example 2F)



Source: Own illustration

[18] See Citro/Michael 1995, p. 161. See Faik 2009, p. 7, too.

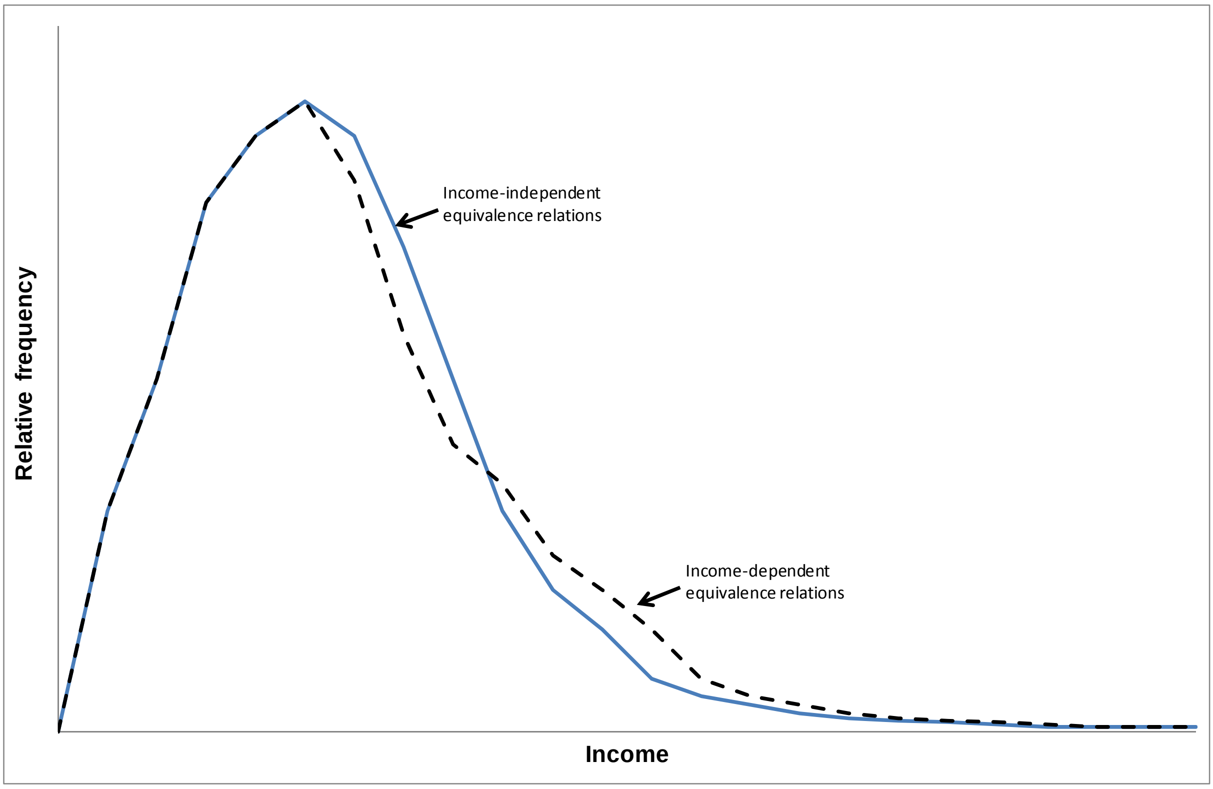
4.2 Income-dependent equivalence relations

There are good reasons for basing distribution analyses on variable equivalence relations. E. g., it might be argued that in the higher income ranges the reference consumption levels (e. g. concerning living) would be fairly high so that the adding of a new household member (e. g. a child) would increase the corresponding costs only marginally, and this would culminate in low *relative* costs in the sense of flat equivalence relations for bigger households in the upper income range compared with the bottom incomes. Another reason for income-dependent scales might be that prices of goods can differ across income groups in that way that members of the upper income classes obtain price advantages.[19]

The estimation of income-dependent equivalence relations is confronted with the initial problem to separate the upper from the bottom range of equivalent incomes. To do this, means that the researcher has to assume concrete equivalence relations for the whole income range as a starting point.[20] To a high degree this is a normative decision. Faik (1995) or Schröder (2004) nevertheless have estimated income-dependent equivalence relations for Germany.[21] Their results indicate indeed lower equivalence relations in the upper income range compared to the bottom income area.

If we use income-dependent equivalence relations with lower scale values for the upper income class, the differences of the equivalent incomes between the lower and the upper income class become bigger than without using income-dependent equivalence relations. So the measured inequality will probably increase.

Figure 7: Income-independent versus income-dependent equivalence relations and their impact on the income distribution



Source: Own illustration

[19] See Schröder 2004, p. 42.
[20] See Faik 1995, pp. 286-287.
[21] Concerning the estimation of income-dependent equivalence relations see further van Hoa 1986, pp. 97-98, or Fiegehen/Lansley/Smith 1977, pp. 105-106.

Figure 7 on the previous page compares the case of income-dependent equivalence relations with the alternative method which uses income-independent equivalence relations. In this ideal-typical illustration it becomes evident that for the sketched distributions, which both are right-skewed, the arithmetic mean as well as the standard deviation arise for the transition from income-independent towards income-dependent equivalence relations. Whether this corresponds with an increase in the measured inequality depends on the strength each of these two increases has. In figure 7 it is assumed – what seems to be realistic – that the relative increase of the standard deviation (in per cent) would be higher than the corresponding increase of the arithmetic mean (in per cent, too) so that as a result the measured inequality would also rise.

Formally, we can show the differences between the case with income-dependent equivalence relations and the case with income-independent equivalence relations by decomposing the numerator (the variance) as well as the denominator (the squared arithmetic mean) of the fraction which the coefficient of variation constitutes. Concretely, the income range is decomposed in two parts, in the bottom and in the upper income area.

Firstly, we refer to the squared arithmetic mean (1 represents the bottom income area, and 2 the upper income area), and we obtain:

$$(27)\quad \mu_*^2 = \left[\frac{n_1 \cdot \mu_{*,1} + n_2 \cdot \mu_{*,2}}{n}\right]^2$$

$$\mu_*^2 = \left[\frac{n_1 \cdot \sum_{g=1}^{G} \frac{\mu_{1,g}}{m_{1,g}} \cdot \frac{n_{1,g}}{n_1} + n_2 \cdot \sum_{g=1}^{G} \frac{\mu_{2,g}}{m_{2,g}} \cdot \frac{n_{2,g}}{n_2}}{n}\right]^2$$

$$\mu_*^2 = \left[\frac{\sum_{g=1}^{G} \frac{\mu_{1,g}}{m_{1,g}} \cdot n_{1,g} + n_2 \cdot \sum_{g=1}^{G} \frac{\mu_{2,g}}{m_{2,g}} \cdot n_{2,g}}{n}\right]^2$$

$$\mu_*^2 = \left[\frac{\sum_{g=1}^{G} \frac{\mu_{1,g}}{m_{1,g}} \cdot n_{1,g} + n_2 \cdot \left(\mu_{2,g=1} \cdot n_{2,g=1} + \sum_{g=2}^{G} \frac{\mu_{2,g}}{a \cdot m_{1,g}} \cdot n_{2,g}\right)}{n}\right]^2 .$$

Secondly, we refer to the variance (1 once more represents the bottom income area, and 2 the upper income area) with the following result:

$$(28)\quad S_*^2 = \frac{n_1 \cdot S_{*,1}^2 + n_2 \cdot S_{*,2}^2}{n} + \frac{n_1 \cdot \left(\mu_{*,1} - \mu_*\right)^2 + n_2 \cdot \left(\mu_{*,2} - \mu_*\right)^2}{n} .$$

Some examples shall illustrate the foregoing connections.

Example 3:

Household size	Household income	Equivalence relation, "first round" and at the bottom = square root of household size (rounded)	Household equivalence income "in the first round" (rounded)	Equivalence relation, at the top= square root of household size * a (rounded)	Household equivalence income "in the second round" (rounded)
1 person	1,000	1.00	1,000	1.00	1,000
2 persons	2,000	1.41	1,414	1.41	1,414
3 persons	2,500	1.73	1,443	1.73 * a	2,500 / (1.73 * a)
4 persons	3,000	2.00	1,500	2.00 * a	3,000 / (2.00 * a)
4 persons	3,500	2.00	1,750	2.00 * a	3,500 / (2.00 * a)

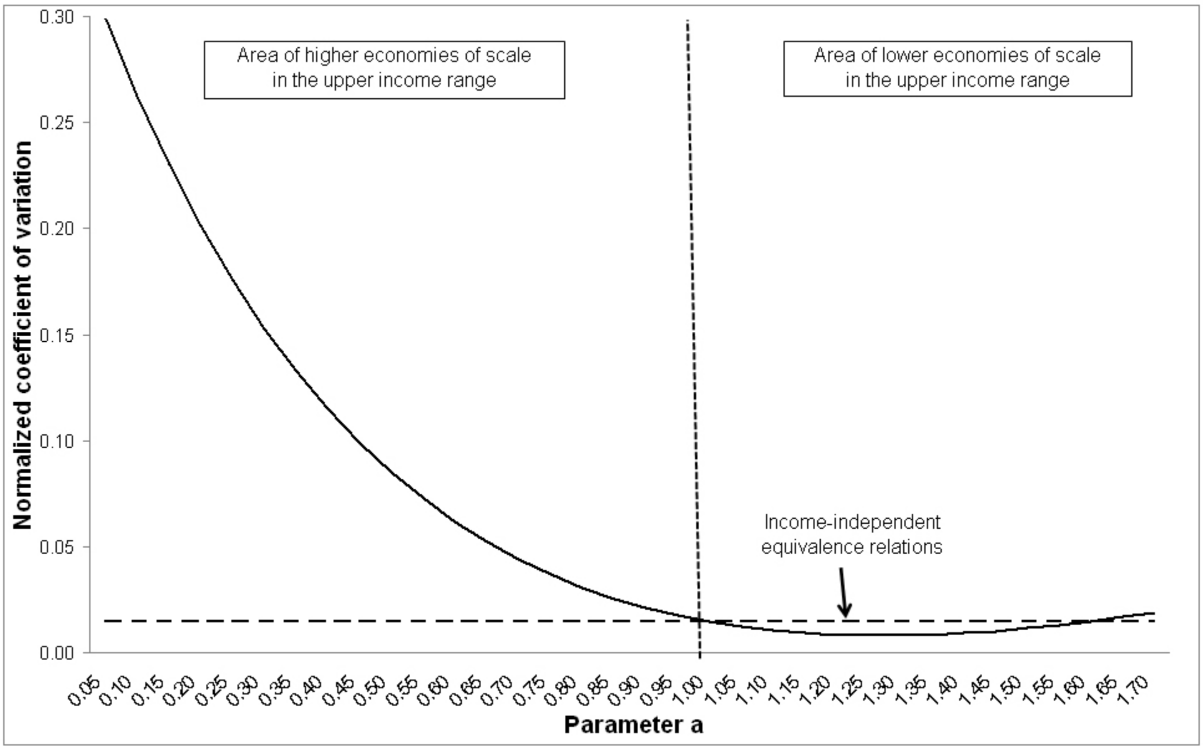
In the first round, a uniform equivalence relation is used in order to differentiate the two income areas from each other. In our example the "square-root scale" was applied. In this case an arithmetic mean for the household equivalence scale in the amount of about 1,422 monetary units (MU) and a standard deviation in the amount of circa 242 MU emerge. These values lead to a normalized coefficient of variation in the amount of approximately 0.0145.

Concerning the resulting equivalence incomes, the first two households belong to the bottom income area because the corresponding values are lower than the computed arithmetic mean. Accordingly, the other three households are members of the upper income area.

Specifying the parameter a as 0.9 in order to compute the equivalence relations in the upper income range, leads to equivalence relations amounting to about 1.56 for the three-person household, and 1.8 for both of the four-person households. The equivalent incomes for these households add up to approximately 1,604 MU (three-person household), 1,667 MU (first four-person household), and 1,944 MU (second four-person household). With these informations we obtain as "new" arithmetic mean value – in the second round – circa 1,526 MU (+ 7 per cent compared with the "old" mean in the first round) and as "new" standard deviation circa 313 MU (+ 29 per cent compared with the "old" standard deviation), so that the "new" normalized coefficient of variations amounts to 0.0211 which is about 46 per cent higher than the "old" CV value. Obviously, the CV value increases, as was expected above, and this is primarily the result of the increasing covariance between the two income areas.

Figure 8 shows the sensitivity effects of different values for the parameter a with respect to CV. It must be emphasized that the area of lower economies of scale in the upper income range at the right of figure 8 has only a mathematical relevance; theoretically, there are no plausible arguments for this area in my eyes. Also, not all values on the left are plausible: So very low values of the parameter a indicate very high economies of scale in the upper income range which would result in (much) smaller equivalence relations for bigger household types in the upper income range compared with smaller household types in the bottom income area; this – at least partly – seems to be not realistic, too.

Figure 8: Income-dependent equivalence relations (example 3) with variation of the parameter a



Source: Own illustration

In example 3 the “variance effect” overcompensated the increasing value of the arithmetic mean at realistic values for the parameter a. In order to show that this effect could notably decline, example 4 was created.

Example 4:

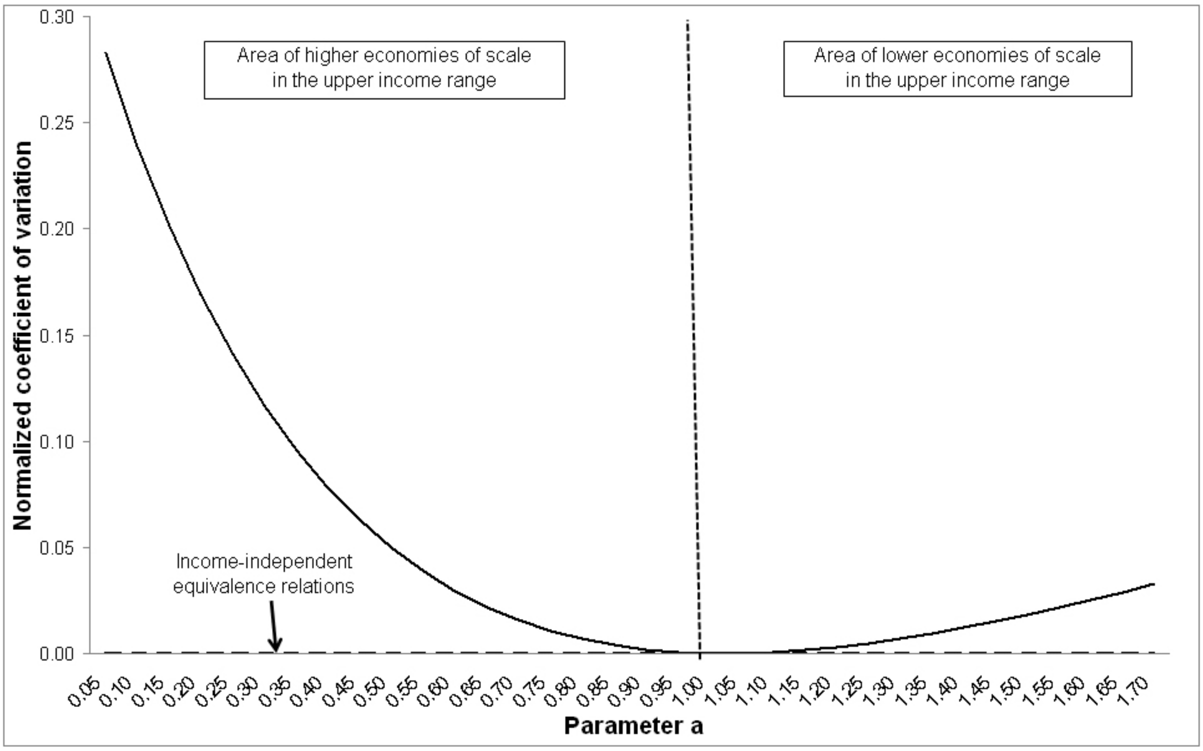
Household size	Household income	Equivalence relation, “first round” and at the bottom = square root of household size (rounded)	Household equivalence income “in the first round” (rounded)	Equivalence relation, at the top= square root of household size * a (rounded)	Household equivalence income “in the second round” (rounded)
1 person	1,100	1.00	1,100	1.00	1,100
2 persons	1,550	1.41	1,096	1.41	1,096
3 persons	1,950	1.73	1,126	1.73 * a	1,950 / (1.73 * a)
4 persons	2,250	2.00	1,125	2.00 * a	2,250 / (2.00 * a)
4 persons	2,250	2.00	1,125	2.00 * a	2,250 / (2.00 * a)

In example 4 the standard deviation of the equivalence incomes “in the first round” shrinks to only circa 14 MU. The value of the arithmetic mean of the equivalence incomes is now about

1,114 MU. Once more, the first two households belong to the bottom income area and the other three households to the upper income area. In the case of income-independent equivalence relations the normalized coefficient of variation only amounts to 0.00007 now.

Depending on the variation of the parameter a, CV varies as it is sketched in figure 9. The outlined shape of the inequality function is similar to the one presented in figure 8 but in the realistic range of values for the parameter a (let us say: from 0.75 to 0.95) the absolute differences are much smaller than in example 3 (as a result of the smaller variance in example 4 compared with example 3): Whereas e. g. in example 3 the difference between the value of CV at a = 0.75 und CV at a = 1.00 amounted to 0.0194, in example 4 the corresponding difference was only 0.0101 (with less CV values in example 4 versus example 3).

Figure 9: Income-dependent equivalence relations (example 4) with variation of the parameter a



Source: Own illustration

5. Concluding remarks

The paper had a rather illustrative character. It dealt primarily with the impact of equivalence relations on inequality. After some methodical considerations this impact was evaluated in some sensitivity computations.

In this context we have discussed – on the exemplary basis of the normalized coefficient of variation (i. e., the GE measure at $\lambda = 2$) – the influence of a set of different equivalence relations with differing economies of scale on inequality. It became evident that the inequality curve (of CV) was u-shaped over the range of (decreasing) economies of scale in the case of a positive correlation between household size and household income. In the opposite case, an extreme negative correlation between household size and household income, the inequality curve (of CV) was continuously sloped positively. So we obtained hints about the sensitivity of distributional results which were caused by different equivalence relations on the one hand and different correlations between household size and household income on the other hand. Careful research on distributional aspects should keep in mind these findings. In my eyes it seems to be necessary at least to think about the concrete selection of equivalence relations in distributional studies (at least in such concerning income or consumption) in a very detailed manner.

In front of such distributional studies the correlation between household size and household income should also be checked (for the reasons stated before). If we e. g. choose for distributional purposes the so-called new and the so-called old OECD equivalence scale, which differently deal with economies of scale,[22] in the case of a negative correlation between the two considered variables another ranking of the measured inequality level might result than in the opposite case of a positive correlation between household size and household income.

Moreover, we have dealt with income-dependent equivalence scales which were separately applied in two income areas, in the bottom and in the upper income range. It could be shown – once more on an exemplary basis – that for realistic parameters the measured inequality was higher than in the case of income-independent equivalence scales. That is – in my eyes – a very interesting aspect, because to my knowledge there have been no distributional applications of income-dependent equivalence relations in Germany until now. But – as was considered above – there are very good reasons for introducing this new element into distributional studies!

All in all, our considerations focused some central distributional aspects which should be absolutely kept in mind in studies concerning the personal income distribution.

References

Buhmann, B., et al. (1988): Income, Well-Being, Poverty, and Equivalence Scales: Sensitivity Estimates Across Ten Countries Using the LIS Database. In: Review of Income and Wealth, 34, pp. 115-142.

Citro, C. F./Michael, R. T. (1995): Measuring Poverty – A New Approach, Washington (D. C.).

Cowell, F. A. (1980): Generalized Entropy and the Measurement of Distributional Change. In: European Economic Review, 13, pp. 147-159.

Faik, J. (1995): Äquivalenzskalen. Theoretische Erörterung, empirische Herleitung und verteilungsbezogene Anwendung für die Bundesrepublik Deutschland, Berlin.

Faik, J. (2007): Elementare Wirtschaftsstatistik, Berlin.

Faik, J. (2008): Ausgewählte Verteilungsbefunde für die Bundesrepublik Deutschland unter besonderer Berücksichtigung der Einkommenslage der älteren Bevölkerung. In: Deutsche Rentenversicherung, # 1-2, pp. 22-39.

Faik, J. (2009): Is the Overall German Personal Income Distribution Constant or Variable over Time? Cross-section Analyses for Germany 1969-2003, FaMa discussion paper #1-2009, Frankfurt am Main.

Fiegehen, G. C./Lansley, P. S./Smith, A. D. (1977): Poverty and Progress in Britain 1953-73, Cambridge.

Hussain, M. A. (2009): The sensitivity of income polarization. Time, length of accounting periods, equivalence scales, and income definitions. In: Journal of Economic Inequality, 7, pp. 207-223.

[22] See e. g. Faik 2009, pp. 7-8.

Jenkins, S. P. (1991): The Measurement of Income Inequality. In: Osberg, L. (ed.): Economic Inequality and Poverty: International Perspective, New York/London, pp. 3-38.

Peichl, A./Pestel, N./Schneider, H. (2009): Demografie und Ungleichheit: Der Einfluss von Veränderungen der Haushaltsstruktur auf die Einkommensverteilung in Deutschland, SOEPpapers on Multidisciplinary Panel Data Research #205, Berlin.

Mookherjee, D./Shorrocks, A. F. (1982): A Decomposition Analysis of the Trend in UK Income Inequality. In: Economic Journal, 92, pp. 886-902.

Rodrigues, C. F. (1993): Measurement and Decomposition of Inequality in Portugal (1980/81 – 1989/90), Documentos de trabalho, No.1/1993, Centro de investigaçâo sobre economia portugesa, Technical University Lisbon.

Shorrocks, A. F. (1980): The Class of Additively Decomposable Inequality Measures. In: Econometrica, 48, pp. 613-625.

Schröder, C. (2004): Variable Income Equivalence Scales. An Empirical Approach, Heidelberg/New York.

van Hoa, T. (1986): Measuring Equivalence Scales. A New System-Wide Method. In: Economics Letters, 20, pp. 95-99.

von Weizsäcker, R. K. (1988): Demographischer Wandel, Staatshaushalt und Einkommensverteilung, Institute for Economics – Department for Public Finance, University Bonn, Discussion Paper No. A – 161, Bonn.

List of symbols

λ	=	Parameter with respect to inequality preferences
μ	=	Arithmetic mean income
μ_*	=	Mean equivalent income
$\mu_{*,1}$	=	Mean equivalent income in the bottom income range
$\mu_{*,2}$	=	Mean equivalent income in the upper income range
$\mu_{1,g}$	=	Mean income in the bottom income range for group g
$\mu_{2,g}$	=	Mean income in the upper income range for group g
μ_g	=	Mean income within group g
μ_{g*}	=	Mean equivalent income in group g
ϕ	=	Geometric mean of incomes
ψ	=	Probability of occurence of an event
θ	=	Parameter for economies of scale
a	=	Factor for economies of scales in the upper income range compared with the bottom income range
CV	=	Normalized coefficient of variation
CV_I	=	Within-group component of CV
CV_{II}	=	Between-group component of CV
G	=	Household size
GE	=	General Entropy index

GE_B	=	Between-group GE inequality measure
GE_g	=	Within-group GE inequality measure
h	=	Entropy for occurrence of an event
H	=	Expectancy value for entropy
i	=	Person i
L_1	=	Standard deviation of the logarithms of income
L_2	=	Logarithmic standard deviation
$m_{1,g}$	=	Equivalence relation for group g in the bottom income range
$m_{2,g}$	=	Equivalence relation for group g in the upper income range
m_g	=	Equivalence relation for group g
MLD	=	Mean logarithmic deviation
n	=	population size
$n_{1,g}$	=	Number of persons of group g in the bottom income range
$n_{2,g}$	=	Number of persons of group g in the upper income range
n_g	=	Number of persons in group g
S^2	=	Variance of incomes
S^2_*	=	Variance of equivalent incomes
$S^2_{*,1}$	=	Variance of equivalent incomes in the bottom income range
$S^2_{*,2}$	=	Variance of equivalent incomes in the upper income range
T	=	Theil’s entropy indicator
T_g	=	Theil’s entropy indicator representing within-group inequality
V	=	Coefficient of variation
v_g	=	Group-specific share of the aggregate income
w_g	=	Population share of group g
Y_i	=	Income of person i
$Y_{i,g}$	=	Income of member i in group g

Folgende FaMa-Diskussionspapiere sind bisher erschienen (Stand: 01. Februar 2010):

FaMa-Diskussionspapier 1/2009: Is the Overall German Personal Income Distribution Constant or Variable over Time? Cross-section Analyses for Germany 1969-2003 *(Jürgen Faik).*

FaMa-Diskussionspapier 2/2009: Alternative Verfahren zur Messung von Armut: Ganzheitliche Methode versus Zerlegungsansatz *(Jürgen Faik)*.

FaMa-Diskussionspapier 3/2009: Zur Frage der Rentenanpassung: Probleme und Lösungsansätze *(Jürgen Faik/Tim Köhler-Rama)*.

FaMa-Diskussionspapier 4/2009: Subjektive und objektive Lebenslagen von Arbeitslosen *(Jürgen Faik/Jens Becker)* [auch erschienen in den SOEPpapers on Multidisciplinary Panel Data Research des Deutschen Instituts für Wirtschaftsforschung (DIW) Berlin als #255 (2009)].

FaMa-Diskussionspapier 5/2009: Wohlstandspolarisierung, Verteilungskonflikte und Ungleichheitswahrnehmungen in Deutschland *(Jürgen Faik/Jens Becker)* [auch erschienen in den SOEPpapers on Multidisciplinary Panel Data Research des Deutschen Instituts für Wirtschaftsforschung (DIW) Berlin als #256 (2009)].

FaMa-Diskussionspapier 1/2010: Methodical Settings in Analyses of the Income Distribution – Some Simple Mathematical Comments *(Jürgen Faik)*.

Alle vorgenannten Diskussionspapiere stehen unter http://www.fama-nfs.de zum Download als PDF-Files zur Verfügung.

NEUE FRANKFURTER SOZIALFORSCHUNG
Jürgen Faik . Hans-Christian Mager | GbR

Nikolausstraße 10
D-65936 Frankfurt/Main
http://www.fama-nfs.de
info@fama-nfs.de
Tel. +49(0)69-34409710
Fax: +49(0)69-34409714